

# Penerapan K-Means Clustering Untuk Pengelompokan Kelas Pada Taman Kanak-Kanak

Dewi Anggraeni<sup>1</sup>, Rizaldi<sup>2,\*</sup>, Guntur Maha Putra<sup>2</sup>

Sistem Informasi, Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Royal, Asahan, Indonesia

Email: <sup>1</sup>anggraeni1987@email.com, <sup>2,\*</sup>rizaldipiliang.rp@email.com, <sup>3</sup>igoenputra@gmail.com

Email Penulis Korespondensi: rizaldipiliang.rp@email.com

Submitted: 21/12/2021; Accepted: 29/12/2021; Published: 31/12/2021

**Abstrak**—Pada dasarnya anak yang usia 5 tahun sudah bisa mengenal huruf, mencocokkan bentuk, mengenal warna kelas dan mudah beradaptasi dengan lingkungan. Namun ada beberapa anak yang memang hanya mempunyai kemampuan membaca saja, namun dikemampuan yang lain tidak ada, begitu pun dengan anak-anak yang lainnya. Berdasarkan hasil pengamatan, peneliti melakukan penelitian penerapan k means clustering untuk mengelompokkan kelas pada taman kanak-kanak. Objek pada penelitian ini adalah anak-anak didik yang sekolah di TK ABA XI. metode yang digunakan untuk pembagian yaitu metode k means clustering, dimana metode ini dapat mengelompokkan data dalam jumlah kecil ataupun dalam jumlah banyak data. Terdapat 2 variabel yaitu variabel kelas safe dan Kelas marwa. Dimana setiap variabel memiliki kriteria yaitu A= Keatifan, B= Mecocokan Gambar, C Mengenal Warna, D= Mengenal Bentuk, E= Menulis. Hasil dari penelitian ini adalah perhitungan k means cluster dalam mengelompokkan kelas pada siswa taman kanak-kanak dapat menjadi acuan bagi guru dalam hal pembagian kelas. Data pengelompokan murid sesuai dengan kapasitas kelas yang diinginkan dan banyaknya kelas bisa berubah sesuai keinginan sekolah.

**Kata Kunci:** K-Means Clustering; Pengelompokan Kelas; Taman Kanak-Kanak

**Abstract**—Basically, children aged 5 years can recognize letters, match shapes, recognize class colors and easily adapt to the environment. However, there are some children who only have the ability to read, but do not have other abilities, as well as other children. Based on the observations, the researchers conducted research on the application of k means clustering to classify classes in kindergarten. The object of this research is students who study in TK ABA XI. The method used for division is the k means clustering method, where this method can group data in small amounts or in large amounts of data. There are 2 variables, namely the variable class\_safa\_and\_Kelas\_marwa. Where each variable has criteria, namely A = Activeness, B = Matching Pictures, C Knowing Color, D = Knowing Shapes, E = Writing. The result of this research is the calculation of k means cluster in grouping classes in kindergarten students can be a reference for teachers in terms of class division. The student grouping data is according to the desired class capacity and the number of classes can change according to the school's wishes.

**Keywords:** K-Means Clustering; Class Grouping; Kindergarten

## 1. PENDAHULUAN

TK ABA X Kisaran adalah jenjang formal pendidikan taman kanak-kanak yang beralamat di Jl. Durian, Kisaran Naga, Kec. Kisaran Tim., Kabupaten Asahan, Sumatera Utara.

Pengelompokkan kelas pada taman kanak-kakan merupakan bagian penting agar menciptakan suasana pembelajaran yang menyenangkan. Perlunya pembagian kelas, agar guru lebih fokus dan terampil pada kemampuan minat dan bakat siswa [1][2][3].

Analisa cluster adalah proses mempartisi sekumpulan objek data ke dalam himpunan bagian. Setiap subset adalah sebuah cluster, sehingga objek-objek dalam sebuah cluster mirip satu sama lain, namun berbeda objek di cluster lain. Pengelompokkan berguna karena dapat mengarah pada penemuan kelompok yang sebelumnya tidak diketahui dalam data. pengelompokan yang dapat digunakan untuk mengatur data yang berukuran besar ke dalam kelompok, dimana setiap data mempunyai karakteristik yang sama. Pengelompokan untuk dapat digunakan untuk menemukan cluster atau subclass disistem.

Algoritma k means dapat mengelompokkan datase yang sama ke dalam cluster yang berbeda. Disini K mendefinisikan jumlah cluster yang telah ditentukan sebelumnya yang perlu dibuat dalam proses, seolah olah k=2 akan ada dua cluster dan untuk k=3 akan ada tiga cluster dan seterusnya.

Algoritma pengelompokkan yang digunakan, baik pada kumpulan data kecil yang berisi lebih sedikit dari beberapa ratus objek data, namun database besar mungkin berisi jutaan atau bahkan miliaran objek, pengelompokan hanya pada sampel dari kumpulan data besar yang diberikan dapat menyebabkan hasil yang bias, oleh karena itu sangat diperlukan algoritma clustering.

Tahapan k means cluster pertama memilih objek secara acak, yang masing-masing awalnya mewakili means cluster atau pusat. Untuk masing-masing objek yang tersisa, sebuah objek dikelompokkan ke cluster yang paling mirip berdasarkan jarak euclidean antara objek dan cluster. [4][5][6].

Algoritma k means secara iteratif meningkatkan variasi dalam cluster, untuk setiap cluster diiterasi sebelumnya. Semua objek kemudian dipindahkan menjadi pusat cluster baru [7][8][9].

Dalam algoritma k means, jumlah kuadrat dari jarak euclidean titik data perwakilan terdekat yang digunakan untuk mengukur fungsi objektif pengelompokan. Persamaan yang untuk mengukur jarak euclidean digunakan yaitu:

$d_{\text{Euclidean}}$

$$j(v_1, v_2) = \sqrt{\sum_{k=1}^N (V_1(k) - V_2(k))^2} \dots \dots [1]$$

Fungsi jarak *euclidean* digunakan untuk menghitung himpunan yang dipilih secara acak dari data sampel, dimana salah satu himpunan mendekati perwakilan dekat dengan cluster dan salah satunya terletak ditengah-tengah cluster.

*Clustering* dapat dikatakan suatu proses pengelompokan objek data yang dibagi menjadi dua bagian yaitu hirarchi dan partisi. [10][11][12]. Perkembangan *K-Means* dalam memodelkan dataset yang mempunyai bentuk khusus dengan memanfaatkan kernel trik, permasalahan yang perlu diperhatikan dalam menggunakan metode *K-Means* termasuk model *clustering* yang berbeda-beda, pemilihan model yang paling tepat untuk dataset yang dianalisa, kegagalan untuk *converge*, pendeteksian *outliers*, bentuk masing-masing cluster dan permasalahan *overlapping* [13][14][15]. Tujuan pengelompokan siswa pada taman kanak-kanak adalah untuk memfasilitasi pertumbuhan dan perkembangan anak secara menyeluruh. Setiap anak mempunyai perkembangan yang berbeda-beda, berdasarkan hal itu pengelompokan dibuat dalam 2 kelas yaitu kelas safa dan kelas marwa.

Hasil penerapan K-Means clustering untuk pengelompokkan kelas pada taman kanak-kanan dapat dijadikan sebagai acuan untuk sekolah dalam mengelompokkan siswa sebelum proses pembelajaran dimulai, dimana pada penelitian hanya menyatakan bahwa setiap siswa masuk kelas safa atau Kelas marwa.

Beberapa penelitian yang menjadi tinjauan pustaka untuk menunjang penelitian ini, diantaranya adalah: Analisis K-Means Clustering pada Data Sepeda Motor[8], Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Maluku[7], Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering[14], dan Penerapan Algoritma K-Means Clustering Untuk Pengelompokkan Penyebaran Diare Di Kabupaten Langkat[13].

Kontribusi dari hasil penelitian ini yaitu memastikan tindakan apa yang harus dilakukan pihak sekolah terhadap siswa dengan kemampuan yang berbeda-beda dan memastikan mana siswa yang masuk ke kelas safa dan mana yang masuk ke kelas marwa yang dimana ke 2 kelas ini mencerminkan 2 kelompok karakter dan 2 kemampuan siswa yang berbeda.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

#### 1. Identifikasi Masalah

Pada tahap ini, peneliti melakukan observasi terlebih dahulu untuk menentukan objek penelitian. Dimana penelitian ini dilakukan disekolah TK ABA Kisaran dan guru menjadi objek pada penelitian ini. Untuk memastikan penelitian ini layak untuk diteruskan diukur dari seberapa besar masalah yang terdapat dan seberapa penting hasil penelitian ini dapat dimanfaatkan.

#### 2. Analisa Data

Data-data yang berhubungan dengan objek yang diteliti dikelompokkan dianalisa satu per satu. Dalam penelitian ini terdapat 2 kelompok data yang di analisa, yaitu nilai standar kelas dan data training. Selanjutnya ada 2 variabel yang digunakan untuk menganalisa data yaitu variabel kelas safa dan kelas marwa. Ada 5 kriteria yang digunakan, yaitu kriteria, kearifan, mencocokkan gambar, mengenal warna, mengenal bentuk, dan menulis yang akan di *convert* ke dalam huruf yaitu A, B, C, D, dan E.

#### 3. Perancangan Sistem

Pada tahap ini merupakan tahap rancangan sistem lanjutan dari analisa data dimana proses-proses yang dilakukan untuk memperoleh hasil yang akurat menggunakan metode *K-Means Clustering* yang diimplementasikan ke dalam sistem yang tujuannya agar pengolahan data lebih efektif dan efisien.

#### 4. Pengujian dan Implementasi

Pada tahapan ini merupakan pengujian sistem dan implementasi, apakah rancangan sistem dengan menerapkan metode *K-Means Clustering* dalam memproses data sudah dapat menjawab masalah yang sudah diidentifikasi sebelumnya atau belum. Jika belum maka perlu dilakukan perbaikan. Kesalahan bisa saja terjadi pada kesalahan input data, kesalahan logika dan aritmatika yang digunakan, dan kesalahan kode-kode dalam bahasa sistem yang digunakan. Selanjutnya, jika sistem aman dari kesalahan maka sistem dapat diimplementasikan.

### 2.2 Tahapan Penerapan Metode K-Means Clustering

Ada 4 tahapan utama yang dilakukan oleh metode K-Means Clustering dalam memproses data, yaitu:

1. Menentukan data standart yang akan di proses menggunakan *K-Means Clustering*
2. Menentukan pusat cluster awal



Hal-hal yang dilakukan pada tahapan ini adalah mengalokasikan data ke dalam cluster dengan acak, kemudian menghitung rata-rata data pada masing-masing *cluster* atau disebut dengan *centroid*.

3. Menghitung jarak cluster

Tahapan ini merupakan tahapan alokasi masing-masing data ke rata-rata terdekat.

4. Memperoleh Hasil Sementara

Apabila masih ada data yang berpindah cluster atau apabila perubahan nilai *centroid*, ada yang di atas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan maka kembali ke step 3.

### 3. HASIL DAN PEMBAHASAN

Metode cluster memisahkan data ke kelompok yang berbeda. Dengan partitioning secara iteratif, k means mampu meminimalkan rata-rata jarak setiap data ke cluster. Pada penelitian ini terdapat 2 variabel yaitu variabel kelas safa dan variabel kelas marwa. Dimana terdapat 5 kriteria, A= kearifan, B=mencocokkan gamabr, C= mengenal warna, D= mengenal bentuk, E= menulis. Dalam pengelompokkan kelas terdapat nilai standar kelas yang harus dipenuhi, berikut adalah data nilai standar kelas:

**Tabel 1.** Nilai Standart Kelas

Siswa	A	B	C	D	E
Kelas safa	75	65	70	80	85
Kelas marwa	90	86	60	75	80

**Tabel 2.** Data Training

Siswa	A	B	C	D	E
Akhdan	80	80	80	80	75
Ihsan	75	80	75	80	70
Al Hassan	85	80	80	85	80
Fahri	90	90	85	80	80
Suryana	75	75	65	75	70
Habibi	80	70	65	80	90
Jihan	70	60	70	80	85
Fahriza	90	70	75	85	90
Hanum	60	75	75	80	75
Zizi	75	75	65	65	70

Berdasarkan data nilai standart dan data training, dapat dilakukan perhitungan dengan menggunakan *K-Means Clustering*, dengan rumus sebagai berikut:

$$d_{\text{Euclidean}}(x,y),(a,b)]=\sqrt{(x - a)^2 + (y - b)^2} \dots(2)$$

1. Menentukan Pusat *Cluster* Awal

Berdasarkan data diatas, diambil secara acak data sebagai pusat *Cluster* 1, yaitu data 8 dan data 9. Data 8 sebagai pusat *cluster* 1, sedangkan data 9 sebagai pusat *cluster* 2.

**Tabel 3.** Cluster Awal

cluster 1	70	60	70	80	85
cluster 2	90	70	75	85	90

2. Menghitung jarak *cluster*

Jarak Cluster 1 ke Cluster 1

$$= [(x,y),(a,b)]=\sqrt{(x - a)^2 + (y - b)^2}$$

$$=\sqrt{(70 - 80)^2 + (60 - 80)^2 + (70 - 80)^2 + (80 - 80)^2 + (85 - 75)^2}= 26,45$$

Jarak Cluster 2 ke cluster 1

$$=\sqrt{(70 - 75)^2 + (60 - 80)^2 + (70 - 75)^2 + (80 - 80)^2 + (85 - 70)^2}= 25,98$$

Jarak Cluster 3 ke cluster 1

$$=\sqrt{(70 - 85)^2 + (60 - 80)^2 + (70 - 80)^2 + (80 - 85)^2 + (85 - 80)^2}= 27,83$$



Jarak Cluster 4 ke cluster 1

$$= \sqrt{(70 - 90)^2 + (60 - 90)^2} + (70 - 85)^2 + (80 - 80)^2 + (85 - 80)^2 = 39,37$$

Jarak Cluster 5 ke cluster 1

$$= \sqrt{(70 - 75)^2 + (60 - 75)^2} + (70 - 65)^2 + (80 - 75)^2 + (85 - 70)^2 = 22,91$$

Jarak cluster 6 ke cluster 1

$$= \sqrt{(70 - 80)^2 + (60 - 70)^2} + (70 - 65)^2 + (80 - 80)^2 + (85 - 90)^2 = 15,811$$

Jarak cluster 7 ke cluster 1

$$= \sqrt{(70 - 70)^2 + (60 - 60)^2} + (70 - 70)^2 + (80 - 80)^2 + (85 - 85)^2 = 15,811$$

Jarak Cluster 8 ke cluster 1

$$= \sqrt{(70 - 90)^2 + (60 - 70)^2} + (70 - 75)^2 + (80 - 85)^2 + (85 - 90)^2 = 23,97$$

Jarak cluster 9 ke cluster 1

$$= \sqrt{(70 - 60)^2 + (60 - 75)^2} + (70 - 75)^2 + (80 - 80)^2 + (85 - 75)^2 = 21,21$$

Jarak cluster r 10 ke cluster 1

$$= \sqrt{(70 - 75)^2 + (60 - 75)^2} + (70 - 65)^2 + (80 - 65)^2 + (85 - 70)^2 = 6,92$$

Jarak cluster 1 ke cluster 2

$$= \sqrt{(90 - 80)^2 + (70 - 80)^2} + (75 - 80)^2 + (85 - 80)^2 + (90 - 75)^2 = 27,38$$

Jarak cluster 2 ke cluster 2

$$= \sqrt{(90 - 75)^2 + (70 - 80)^2} + (75 - 75)^2 + (85 - 80)^2 + (90 - 70)^2 = 27,38$$

Jarak cluster 3 ke cluster 2

$$= \sqrt{(90 - 85)^2 + (70 - 80)^2} + (75 - 80)^2 + (85 - 80)^2 + (90 - 80)^2 = 15,811$$

Jarak cluster 4 ke cluster 2

$$= \sqrt{(90 - 90)^2 + (70 - 90)^2} + (75 - 85)^2 + (85 - 80)^2 + (90 - 80)^2 = 51,96$$

Jarak cluster 5 ke cluster 2

$$= \sqrt{(90 - 75)^2 + (70 - 75)^2} + (75 - 65)^2 + (85 - 75)^2 + (90 - 70)^2 = 70,53$$

Jarak cluster 6 ke cluster 2

$$= \sqrt{(90 - 80)^2 + (70 - 70)^2} + (75 - 65)^2 + (85 - 80)^2 + (90 - 90)^2 = 77,32$$

Jarak cluster 7 ke cluster 2

$$= \sqrt{(90 - 70)^2 + (70 - 60)^2} + (75 - 70)^2 + (85 - 80)^2 + (90 - 85)^2 = 36,05$$

Jarak cluster 8 ke cluster 2

$$= \sqrt{(90 - 90)^2 + (70 - 70)^2} + (75 - 75)^2 + (85 - 85)^2 + (90 - 90)^2 = 70,06$$

Jarak cluster 9 ke cluster 2

$$= \sqrt{(90 - 60)^2 + (70 - 75)^2} + (75 - 75)^2 + (85 - 80)^2 + (90 - 75)^2 = 71,46$$

Jarak cluster 10 ke cluster 2

$$= \sqrt{(90 - 75)^2 + (70 - 75)^2} + (75 - 65)^2 + (85 - 65)^2 + (90 - 70)^2 = 71,46$$

Dari hasil perhitungan jarak data yang diambil dengan cluster dapat dilihat pada tabel dibawah ini :

**Tabel 4.** Hasil Cluster Sementara

Siswa	Centroid		Kelas
	C1	C2	
Akhdan	26,45	27,386	safa
Ihsan	25,98	27,238	safa

Siswa	Centroid		Kelas
	C1	C2	
Al Hassan	27,88	15,811	marwa
Fahri	38,37	51,811	safa
Suryana	22,91	70,98	safa
Habibi	15,811	77,323	safa
Jihan	26,7	36,005	marwa
Fahriza	23,97	70,062	safa
Hanum	21,213	71,46	marwa
Zizi	26,74	67,478	safa

#### 4. KESIMPULAN

Hasil ini penelitian ini merupakan pengelompokkan kelas pada siswa taman kanak-kanak yang tujuannya adalah memastikan tindakan apa yang harus dilakukan pihak sekolah terhadap siswa dengan kemampuan yang berbeda-beda dan memastikan mana siswa yang masuk ke kelas safa dan mana yang masuk ke kelas marwa yang dimana ke 2 kelas ini mencerminkan 2 kelompok karakter dan 2 kemampuan siswa yang berbeda.

#### REFERENCES

- [1] L. A. Shofiana and L. A. Muharom, "Aplikasi K-Means Untuk Pengelompokan Siswa Pendidikan Anak Usia Dini ( Paud ) Berdasarkan Data Siswa Di Kecamatan Ledokombo," no. 1110651055, pp. 0-2, 2009.
- [2] Asroni and R. Adrian, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *J. Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76-82, 2015.
- [3] S. S. Nagari and L. Inayati, "Implementation of Clustering Using K-Means Method To Determine Nutritional Status," *J. Biometrika dan Kependud.*, vol. 9, no. 1, p. 62, 2020, doi: 10.20473/jbk.v9i1.2020.62-68.
- [4] S. S. Helma, R. R. R., and E. Normala, "Clustering pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma K - Means," no. November, pp. 131-137, 2019.
- [5] R. A. Indraputra and R. Fitriana, "K-Means Clustering Data COVID-19," *J. Tek. Ind.*, vol. 10, no. 3, p. 3, 2020.
- [6] M. Robani and A. Widodo, "Algoritma K-Means Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia," *J. Sist. Inf. Bisnis*, vol. 6, no. 2, p. 164, 2016, doi: 10.21456/vol6iss2pp164-176.
- [7] M. W. Talakua, Z. A. Leleury, and A. W. Taluta, "Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 11, no. 2, pp. 119-128, 2017, doi: 10.30598/barekengvol11iss2pp119-128.
- [8] R. K. Dinata, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," vol. 5, no. 1, 2020.
- [9] T. N. P. Tesa Nur Padilah, Cepy Sukmayadi, "Clustering Fasilitas Kesehatan Berdasarkan Kecamatan," vol. 8, no. 1, pp. 83-92, 2021.
- [10] F. M. Yuma and R. Rizaldi, "Sistem Pendukung Keputusan Dalam Pemilihan Jurusan Di Perguruan Tinggi Dengan Metode Profile Matching," *Semin. Nas. R.*, vol. 1, no. 1, pp. 181-184, 2018.
- [11] W. Dhuhita, "Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita," *J. Inform. Darmajaya*, vol. 15, no. 2, pp. 160-174, 2015.
- [12] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.
- [13] F. Nasari and C. J. M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat," *CogITO Smart J.*, vol. 2, no. 2, p. 108, 2016, doi: 10.31154/cogito.v2i2.19.108-119.
- [14] S. Handoko, E. Tri, and E. Handayani, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING," vol. 25, no. 1.
- [15] F. Nur, M. Zarlis, and B. B. Nasution, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolahmenengah Kejuruan Untuk Clustering Jurusan," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 1, no. 2, pp. 100-105, 2017, doi: 10.30743/infotekjar.v1i2.70.